

Q1

MEASUREMENT OF SCHOLASTIC ATTAINMENT WITH STANDARDIZED TESTS

Carson Y. Nolan and Carl J. Davis

Use of standardized achievement tests for evaluating the scholastic attainment of children in the elementary grades through high school is a widespread practice. For example, Cronbach (1960) lists ten separate batteries of achievement tests as examples of those available for this purpose. In 1918 Hayes reported to the American Association of Instructors of the Blind on the use of standardized tests for the seeing in the areas of reading, spelling, arithmetic and language ability in three schools for the blind. Lowenfeld (1963) summarizes attempts to use standardized achievement tests with the visually-impaired in the United States since that early date. Among the tests he mentions in use for this purpose can be found the Stanford Achievement Tests, Gray Oral Reading Check Lists, Metropolitan Achievement Tests, Sones-Harry High School Achievement Test, Myers-Ruch High School Progress Test, as well as tests offered by the College Entrance Examination Board.

This use of standardized tests has not been without its problems. Hayes, addressing the first I.C.E.B.Y. in 1952 stated: "In adapting such tests for use with blind pupils two major considerations were kept in mind: (1) Is there anything in the content of the tests which is unfair to pupils without vision? and (2) How can we be sure that the blind are not handicapped in the mechanics of reading and marking the test material?" Nolan (1962) discussed these and other areas of concern in the following quoted paragraphs.

Problems in Test Use

"Throughout the story of the use of standardized achievement tests with the visually handicapped, there is an underlying theme of frustration and doubt. Questions arising as to the meaning of individual test scores obtained from adapted tests have resulted in constant frustration of efforts to apply them to the solution of common educational problems. Why should this be the case? The answer lies in the fundamental nature of achievement tests as implied in the term 'standardized'.

Much of the usefulness of standardized achievement tests results from the fact that they provide a common frame of reference for evaluating scholastic achievement of children everywhere. This is accomplished by holding constant a number of variables in the testing situation. Held constant are the test content (the actual test questions), the method by which the test is administered, the length of time allowed for each part of the test, and the method by which the test is scored. Meaning is given to scores on achievement tests by expressing them in terms of the average achievement of a national sample of school children for each month of the school year at each grade level.

Thus, standardized achievement tests are valuable because we are able to assess each child's achievement under approximately identical conditions. Tests give us a common measuring stick for comparing achievement of individual children, achievement of classes within the same school, and achievement between educational programs. This is based upon the assumption that all children are exposed to the same curriculum under equal conditions for learning. The reference marks on our educational measuring stick are established empirically by the number of test questions that will be answered correctly by the average child at each grade level for each month of the school year. Since these have been established under standard conditions of testing, if we change the testing conditions, the previously established reference marks are no longer true or valid. Adaptation of standardized tests for use with visually handicapped children almost invariably involves changes in the standard conditions of testing which result in loss of validity or meaning for the scores obtained. Changes commonly made in standard testing conditions include changes in the time allotted for tests, changes in the methods for administration of tests, and changes in test content. Each of these changes influences the meaning of scores obtained by test use."

The Problem of Time Limits

This problem has been described by Nolan (1962) as follows:

"Because of the slowness of Braille reading, it is necessary to change the standard time limits of tests for visually handicapped children. To the present,

a somewhat arbitrary ratio of 2.5/1 was used as an adjustment factor. This ratio is based on the comparative reading speeds for Grade 1½ Braille and print reading as established many years ago (Hayes, 1937). Since tests are currently published in Grade 2 Braille, this ratio may no longer be appropriate. In addition, the use of any single ratio for adjusting test scores assumes that the ratio between Braille and print reading speed is the same at all grade levels and for all forms of material commonly presented in achievement tests. This, however, has never been established as a fact. The effect of time limits, particularly for the lower grades at any test level, is to restrict the amount of opportunity that children have to attempt to answer test items. We do not know, at present, how the time extensions made for Braille reading effect the opportunity of visually handicapped children to attempt test items for the various types of tests. Therefore, we have a source of ambiguity in interpreting a Braille test score."

An identical problem exists in the case of large type tests.

The solution to the problem of time limits appears to be more logical than empirical. In the directions for the most recent edition of the Stanford Achievement Tests, the authors (Kelley, et al., 1964) state, "Although almost all tests are time-limit tests, the time limits are provided as matters of administrative convenience rather than for the purpose of placing any premium upon speed of work. The time limits in all cases are generous and calculated to give practically all pupils sufficient time to attempt all questions which they are capable of answering correctly. The tests, therefore, are fundamentally power tests, and not speed tests." Such statements appear in the manuals of most academic achievement tests, consequently, it seems advisable to allow visually handicapped students sufficient opportunity to attempt as many test items as they are capable of without respect to time.

Problems of Administrative Methods

It is sometimes the practice to administer orally achievement tests to visually handicapped students. Under these conditions, the examiner reads aloud

the questions and the possible answers. The students mark their responses either in regular test booklets, in abbreviated test booklets containing only the answers to questions, or on special answer sheets listing only the letter or the response position identifying the answer. The purposes of this practice vary, but are usually concerned with decreasing testing time or providing measures for students whose reading ability is low or nil. Newly blinded students would be a case in point. Under these conditions, no time limits are imposed and every child has an opportunity to attempt to answer every question.

This practice results in inflation of test scores, particularly for the lower grades in each test level, when results of oral administration are compared with those of braille tests given under the former conditions of time limits which were 250% greater than those for the seeing (Davis & Nolan, 1961). Part of these differences appear to result from unequal opportunity to attempt all test items. Removing the time limits for the written tests would equate for this factor. It does seem, however, that oral testing would diminish that part of test score variance that is attributable to differences in reading ability among students and consequently, result in different score distributions than those obtained using written tests. This question is open to empirical exploration. An additional variable is due to differences in rate and quality of reading that exist among classroom teachers. It does seem desirable to use oral methods of test administration only as a last resort.

Problems of change in Test Content

Many achievement tests contain items that are not suitable for visually handicapped students who read braille. These are test items that are expressed in picture, chart or graph form and which cannot be expressed tactually or require behavior not possible by blind children. In such cases, it is necessary to omit these items or to change their form substantially. Either of these solutions violates the standard conditions of the test. If we omit a test item, we change the number of increments in our measuring instrument and the meaning of the norm

scores for the test is no longer the same. If a test item is changed, while we still have the same number of increments of measurement, the degree of difficulty of the increments likely has been changed with a subsequent loss of meaning. In the past, the effects of such changes in test content upon test scores has been largely ignored. Recently, however, methods have been put to use which enable us to overcome these difficulties.

Adjusting Test Norms for Changes in Content

Examples of such adjustments can be found in the recent braille adaptations of the new edition of the Stanford Achievement Tests and of Form B, Sequential Tests of Educational Progress (STEP). However, the procedures employed in these two instances were radically different.

In the case of the recent editions (W, X, Y, & Z) of the Stanford Achievement Tests, we were fortunate enough to be able to start adaptation during the norming phase for Form X. Early contacts made by Davis of Perkins School with Harcourt Brace & World, the publishers, paved the way. We were able to obtain the copy for the final revision of Form X before these tests were published. Editing by the American Printing House revealed that certain test items had to be omitted from the social studies, science and arithmetic subtests of most batteries. The item numbers for the questions omitted in each subtest were sent to the test publisher who recomputed the sighted norms using only those items to be included in the braille edition. This gave us, then, norms from the original standardization population based only on the items to be used for the blind.

To adapt successive forms of the test (W, Y, & Z), it is necessary only to follow common procedures, and to equate each of these forms to Form X. To accomplish this for Form W, the subtests of Form X and W from which items were dropped for the blind were given to a sample of about 300 sighted students in each grade covered by the tests. Half the students took Form W first and half took Form X first with a two week period separating test administrations. After the tests were administered, the subtests to be equated were scored only for those items to be used

in the Braille edition. Frequency distributions of test raw scores were then made combining all grades taking each subtest within each battery of Form X and Form W. This resulted in one frequency distribution for each subtest within each battery for each form. The two raw score distributions for Forms X and W on each subtest were then plotted on a normal percentile chart and smooth curves fitted to the points plotted. For each Form W raw score, it was possible to read the corresponding raw score for Form X. Since we already had the norm scores or grade scores for Form X, it was a simple matter to assign this same norm score to the corresponding raw score for Form W.

Adjusting the norms of subtests of the STEP tests was an operation of much greater technical difficulty, since it was not possible to use the original normative data. A general description of the procedure can be found in Trismen (1967) and a more technical account in Trismen (1965). Briefly stated, the tests were edited and items to be omitted identified. Next, groups of 300 blind students at each appropriate grade were administered the shorter brailled STEP subtests under conditions of unlimited time. After test score statistics for each blind group had been computed, it was possible to identify, from among STEP data on file at Educational Testing Service, sighted groups of approximately the same achievement levels as those demonstrated by the blind subjects. Since the items included in the braille tests were also included in the regular print tests, data were available to compare both sighted and blind groups on parallel forms of the same shortened test. This made it possible to equate the distribution of scores for the blind to the distribution of the sighted on the shortened form of each subtest. Next, it was possible, through knowledge of the relationship between the distributions for the sighted on the shortened form and regular form of each subtest, to estimate the standing of a blind student with respect to the sighted population's performance on the total subtest. As a result, conversion tables could be constructed which would enable us to interpret the blind student's score using the regular sighted norms for the full test.

Summary

Three problems have been discussed as critical to the use of standardized achievement tests with the visually handicapped. These are the problems of changes in time allotted for tests, changes in the methods for administration of tests, and changes in test content. All are pertinent to adaptation of tests to braille form, while only the first is a problem for large type adaptations.

The problem of time limits was discussed in the context of test authors definition of their tests. Achievement tests are almost invariably defined as power tests with time allotments set for administrative convenience only. Therefore, it appears reasonable to allot to visually handicapped students as much time as they require.

Variation in ability to read among the sighted normative sample undoubtedly is responsible for a significant part of the total variation in the normative score distributions. While the relative contribution of reading to test score variability for the normally sighted as compared with the visually handicapped group is unknown, it does seem desirable to retain the written form of test administration whenever possible. Oral administration of tests should be only a last resort.

When test items must be omitted from the braille forms of tests, it is imperative that test norms also be adapted. This can be accomplished through the use of statistical equating procedures, two of which are mentioned. This process becomes most expeditious if it can be undertaken as part of initial test construction.

References

- Cronbach, L. J. Essentials of psychological testing. (2nd ed.) New York: Harper & Row, 1960, p.382-384
- Davis, C.J. & Nolan, C.Y. A comparison of the oral and written methods of administering achievement tests. International Journal for the Education of the Blind, 1961, 10, 80-82.
- Hayes, S.P. Standard Tests in Elementary Subjects in Schools for the Blind, American Association of Instructors of the Blind, 1918, Colorado Springs, Colo.
- Hayes, S.P. The measurement of educational achievement in schools for the blind. Teachers Forum, 1937, 9, 82-90.
- Hayes, S.P. The Needs of the Blind Child in the Field of Educational Psychology, Proceedings of I.C.E.B.Y., Bussum, Holland, 1952.
- Kelley, T.L., Madden, R., Gardner, E.F., & Rudman, H.C. Stanford Achievement Test: Directions for administering, Advanced Battery. New York: Harcourt, Brace, & World, 1964, p.3.
- Lowenfeld, B. Psychological problems of children with impaired vision. In Cruickshank, W.M. (Ed.) Psychology of exceptional children and youth. Englewood Cliffs, N.J.: Prentice Hall, 1963, p. 242-243.
- Nolan, C.Y. Evaluating the scholastic achievement of visually handicapped children. Exceptional Children, 1962, 28, 493-496.
- Trismen, D.A. Equating braille forms of the Sequential Tests of Educational Progress. Statistical Report 65-47. Princeton: Educational Testing Service, 1965.
- Trismen, D.A. Equating braille forms of the Sequential Tests of Educational Progress. Exceptional Children, 1967, 33, 419-424.

THEORY OF THE EARTH AND ITS HISTORY

CHAPTER I

OF THE ORIGIN AND DEVELOPMENT OF THE EARTH

SECTION I

OF THE ORIGIN OF THE EARTH

THE EARTH IS BELIEVED TO HAVE BEEN FORMED BY THE ACCUMULATION OF MATTER

FROM A GASEOUS STATE INTO A SOLID STATE

THE FIRST STATE OF THE EARTH WAS A GASEOUS STATE

OF MATTER

THE GASES WERE HEATED BY THE SUN'S RAYS

AND BECAME CONDENSED INTO LIQUID

AND SOLID FORMS

THE SOLID FORMS WERE FORMED BY THE

COOLING OF THE LIQUIDS

THE LIQUIDS WERE FORMED BY THE

CONDENSATION OF THE GASES

THE GASES WERE FORMED BY THE

HEATING OF THE MATTER

THE MATTER WAS HEATED BY

THE SUN'S RAYS

THE MATTER WAS

CONDENSED INTO

SOLID FORMS